UNITED STATES DEPARTMENT OF COMMERCE
Bureau of the Census
Washington, DC 20233-0001

September 21, 2000

DSSD CENSUS 2000 PROCEDURES AND OPERATIONS MEMORANDUM SERIES B-6

MEMORANDUM FOR Howard Hogan
Chief, Decennial Statistical Studies Division

From: David C. Whitford DCW
Assistant Division Chief for Statistical Program Management
Decennial Statistical Studies Division

Prepared by: Danny R. Childers DRC and Rosemary Byrne RLB
Decennial Statistical Studies Division

Subject: Accuracy and Coverage Evaluation Survey:
Person Matching and Follow-up Results (Prototype)

The attached document is a prototype of the report that we will prepare, per your request, following completion of applicable Accuracy and Coverage Evaluation Survey (A.C.E.) operations. The completed report is intended to aid the Executive Steering Committee on A.C.E. Policy (ESCAP) in its recommendation regarding the release of the statistically corrected data or the data without statistical correction as the P.L. 94-171 data. This report, together with other reports, will assess the operations and results of both the initial Census and the A.C.E. Both sets of assessments will be available to the ESCAP to aid the Committee in reaching its recommendation regarding the use of the statistically corrected data.

The attached prototype contains both empty table shells and a description of textual analysis that will assess specific aspects of the applicable operations. This report focuses on person matching and follow-up results.

It is important to note that the conduct of the operations may lead us to modify the attached format by including additional information. It is also likely that descriptions and definitions will be enhanced or the data items could undergo revision. Conversely, we may conclude, for a variety of reasons, that some of the information set forth in the attached prototype may not be available. The attached document sets forth our conclusions prior to completion of the A.C.E. about what information would properly inform the ESCAP on this subject, but is subject to modification.

Accuracy and Coverage Evaluation 2000: Person Matching and Follow-up Results

prepared by Danny R. Childers and Rosemary Byrne

Introduction.

This memorandum documents the results of the person matching and follow-up operations. The person matching results are presented for both before and after follow-up matching. Quality assurance results are presented for each step in the clerical matching and follow-up interviewing operations.

Matching refers to the determination of whether an individual enumerated in the A.C.E. is the same person as an individual enumerated in the census. Because errors in matching can significantly affect undercount estimates, highly accurate matching is an important component of the A.C.E. methodology. Although neither Secretary Mosbacher nor the Committee on Adjustment of Postcensal Estimates (CAPE) identified matching error as a significant problem with the 1990 PES, the Census Bureau has made significant improvements to the matching process in the 2000 A.C.E. design, and matching error is expected to be even lower in Census 2000 than in 1990.

Several deficiencies in the 1990 Post Enumeration Survey (PES) design and matching operations prompted improvements in the 2000 A.C.E. matching operation. One problem in 1990 was the misreporting of census day address, with an estimated 0.7 percent of the P-sample being erroneously reported as nonmovers (West 1991). The 2000 A.C.E. improves on 1990 PES in several ways. The CAPI instrument should improve the quality of the reporting of mover status because it is a more automated process. Also, unlike in 1990 where each inmover household (those that moved into PES block clusters after census day) had to be matched to a census day address, which was usually outside the cluster, in the 2000 A.C.E. the reconstructed census day household is matched to the census enumerations in the sample block cluster. The census day household consists of nonmovers and outmovers. The nonmovers live in the housing unit at the time of the interview and on census day. The outmovers lived in the housing unit on census day, but moved before the A.C.E. interview. For clusters with high rates of A.C.E. housing unit nonmatch and census geocoding error, the search area was extended in Census 2000 to the surrounding blocks. Note that the unresolved match codes due to incomplete mover address will not exist in A.C.E. 2000 because there is no inmover matching operation.

A study of clerical error in the 1990 PES found error in coding matches (Davis 1991a) and erroneous enumerations (Davis 1991b). In 1990, codes were entered into a computer system, but the actual matching and duplicate searches were done using paper. In 2000 A.C.E., we expect the matching to be better controlled and more efficient because the clerical matching and quality assurance are fully automated. The automated interactive system will not prevent all matching error, but should reduce the chances for error significantly. Examples of the improvements in

coding are:

- Electronic filtering allows searching based on first name, last name, characteristics, and addresses. For example, the system allows searching for all people named George, all people whose last name begins with a H, all people on Elm Street, or everyone between 30 and 40.
- Only particular codes that fit the situation are allowed. For example, only P-sample nonmatch codes can be assigned a P-sample nonmatch after follow-up code.
- The electronic searches for duplicates will reduce the tedious searching through paper lists of census people. The searching in 1990 was limited to printouts in two sorts: last name and household by address. In 2000, the clerks will have the filtering on name, characteristics, and address to help identify duplicates.
- The system monitors whether the matcher has completed all the necessary searches such as looking for duplicates.
- There are built in edits to check for consistent coding. For example, codes that apply to a household are assigned to all people in the household, such as a geographic code.
- The system automatically assigns certain codes, minimizing coding error.
- A code to indicate that the case needs review at the next level of matching is available to the clerical matchers. This code allows them to flag unusual cases to be done by a person with more experience.
- All quality assurance for the clerical matching is automated. The quality assurance cannot be skipped in 2000.
- Clerical matching is centralized at the National Processing Center instead of different groups of matchers in the seven processing offices, as was done in 1990. Forty six Technicians were hired in September 1999 and have been thoroughly trained in the design of the A.C.E. and matching of people and housing units. These Technicians will perform the quality assurance for the clerical matchers. Additionally, ten Analysts are our most experienced matchers. The Analysts will do the quality assurance for the Technicians and handle the most difficult cases.

Person Matching Results

The P-sample people that go into matching are the nonmovers, the outmovers, and the people with unresolved residence status from the A.C.E. interviewing. These census day residents should have been enumerated in the census. The inmovers are not included in the P-sample for matching. The E-sample is the corresponding census people in the sample block cluster.

The P-sample and census people within the sample block cluster were computer matched followed by a clerical review. The matching steps are:

- The matching is first between the P-sample and the E-sample.
- The remaining not matched P-sample people are searched among the non E-sample people in the sample block cluster, which includes the people enumerated in group

quarters.

- P-sample and E-sample duplicates are identified clerically.
- In clusters selected for targeted extended search, the surrounding blocks are searched clerically for P-sample matches and possible matches and census people geocoded in the surrounding blocks are coded as correctly enumerated.

P-sample and E-sample nonmatches are sent for a follow-up interview. The results of the interview are clerically recorded in the matching software. The results of before and after follow-up coding are displayed to monitor the data in order to identify anomalies in the A.C.E. and census data.

Before Follow-up Results

The first two tables contain the results of before follow-up matching for the P-sample and the E-sample. For details of these codes, see Childers (2000). These before follow-up matching results are from unweighted data from the fifty states and the District of Columbia. These tables do not include the before follow-up matching results in Puerto Rico. The P-sample codes are grouped into

- Matched
- Not matched
- Possible match
- Unresolved match status
- Removed from the P-sample

Matched - The P-sample person is found in the census.

Not Matched - The P-sample person is not found in the census. A follow-up interview is conducted for

- partial household nonmatches
- whole households of conflicting household members (i.e., whole households of P-sample and census nonmatches)[1]
- other whole household nonmatches where the P-sample interview was conducted with a nonhousehold member[2]

Possible Match - The P-sample person may be a match to the census person. A follow-up interview is needed to determine if the two names refer to the same person.

---

[1] These cases have been called the Smith/Jones cases in the past.

[2] No follow-up interview is conducted when there are whole households of P-sample nonmatches from interviews with household members in a housing unit that did not match in the housing unit operation or matched to a housing unit containing no data defined people.

Unresolved Match Status - The only category of unresolved before follow-up is insufficient information for matching and follow-up for the P-sample person.

Removed from the P-sample - The only category of removed from the P-sample in the before follow-up matching are the P-sample people coded as duplicates. The P-sample duplicates are removed because they are listed more than once.

The E-sample codes are grouped into

- Correctly enumerated
- Erroneously enumerated
- Not matched and needing a follow-up interview
- Possible match
- Unresolved

Correctly enumerated - At this point, the only correctly enumerated people are the ones matching the P-sample.

Erroneously enumerated - The categories during before follow-up are fictitious people, duplicates, insufficient information for matching and follow-up, and geocoding errors.

- The fictitious people are ones where we found notes on the census image identifying the person as not a real person such as a dog or other pet.
- The E-sample people enumerated more than once are coded as duplicates.
- The E-sample people with insufficient information for matching and follow-up are ones who are data defined, but do not contain full name and at least two characteristics.[3]
- Census people in housing units identified as geocoding errors[4] during the housing unit follow-up are coded as erroneously enumerated because of geocoding error.

Unresolved enumeration status - In before follow-up matching, the unresolved category only includes the census housing units needing targeted extended search field work that was not done.

E-sample nonmatches - All E-sample people who do not match to the P-sample are sent for a follow-up interview.

E-sample possible matches - E-sample people who were coded as possible matches are followed up to determine whether they are, in fact, matches.

---

[3] This is the same rule that was used in the 1990 PES. There must be enough information about the person to have a chance at locating the person for a follow-up interview before the person is allowed into the matching process. See Childers (2000).

[4] A geocoding error is an error in assigning the housing unit to the correct location.

Table 1: National P-Sample Before Follow-up Matching		
P-sample Match Status	Unweighted People	Percent
Matched		
Not Matched		
Possible Match		
Unresolved		
Removed		
Total		

Table 2: National E-Sample Before Follow-up Matching		
E-sample Enumeration Status	Unweighted People	Percent
Correctly Enumerated		
Erroneously Enumerated		
Unresolved		
Not Matched		
Possible Match		
Total		

Preliminary Census Day Interview Outcome

The preliminary interview outcome codes in the next table identify interviews and noninterviews in occupied housing units, vacant housing units, and housing units that are removed from the P-sample. The interview outcomes in "Accuracy and Coverage Evaluation Survey: Person Interviewing", (Feindt 2000) are the interview outcomes for interview day. The interview outcomes described in this section are census day interview outcomes after data editing, which converts whole households of census day residents with insufficient information for matching to noninterviews and whole households of census day residents who should not have been counted at the housing unit on census day to vacant housing units.

Interviews -

- Complete interviews - interviews conducted with a household member.
- Proxy interviews - interviews conducted with someone outside the household
- Sufficient partial interviews - interviews with household members or proxies that

collected not all information, but enough information to be considered complete interviews.

Noninterviews -

- Field noninterview
- Whole households of people with insufficient information for matching and follow-up

Vacant on Census Day -

- Housing units identified as vacant on census day by the interviewer
- Whole households of people who should have been counted elsewhere on census day (i.e., whole household nonresidents)

Not a Housing Unit on Census Day -

- The housing units identified during the person interview as not a housing unit on census day are removed from the P-sample.

Table 3: Preliminary Census Day Interviewing Outcome for P-sample Housing Units

Outcome Code	Unweighted Housing Units	Percent
Complete interview with a household member		
Complete interview with a proxy respondent		
Sufficient partial interview		
Field noninterview		
All people have insufficient information for matching and follow-up		
Vacant on census day		
No census day residents		
Not a housing unit on census day		
Total		

The noninterview rates in the next three tables are for occupied housing units. The interviewed housing units and the noninterviewed housing units are added together yielding the total number

of occupied housing units.

$$\text{Noninterview Rates for Occupied Housing Units} = \frac{\text{Noninterviews}}{\text{Interviews + Noninterviews}}$$

These data are unweighted sample data. Table 4a contains the percent noninterview for the twelve regional offices and the nation. Tables 4b and 4c contain the P-sample noninterview rates for the census regions and the different types of enumeration areas.

Table 4a: P-sample Preliminary Noninterview Rates in Occupied Housing Units by Census Regional Office
(Unweighted Data)

Census Regional Office	Percent Noninterview
Boston	
New York	
Philadelphia	
Detroit	
Chicago	
Kansas City	
Seattle	
Charlotte	
Atlanta	
Dallas	
Denver	
Los Angeles	
Total	

Table 4b: P-sample Preliminary Noninterview Rates in Occupied Housing Units by Census Region (Unweighted Data)	
Census Region	Percent Noninterview
Northeast	
South	
Midwest	
West	

The Census Bureau defines type of enumeration area (TEA) codes at the census collection block level. Each block must have a TEA code, and no block may have more than one TEA code.

- Mailout/Mailback - The USPS delivered the census questionnaires to city-style addresses by mail and the respondent returned the questionnaire by mail or they were visited during nonresponse follow-up.
- Update/Leave - Address lists were compiled by address listing. The questionnaires were delivered by enumerators who updated the address lists when new addresses were discovered. The respondent returned the completed questionnaire by mail or they were visited during nonresponse follow-up.
- List/Enumerate - Enumerators visited these remote and sparsely populated areas listing the housing units and enumerating the residents.
- Rural Update/Enumerate - The enumerators began with address lists (for previously update/leave areas), updated their lists of addresses, and completed census questionnaires for the residents.
- Urban Update/Leave - The enumerators began with address lists (for previously mail out/mail back areas), updated their lists of addresses, and left census questionnaires for the residents to complete and return by mail or they were visited during nonresponse follow-up.
- Urban Update/Enumerate - The enumerators began with address lists (for previously mail out/mail back areas), updated their lists of addresses, and completed census questionnaires for the residents.
- Additions to Address Listing Universe of Blocks - Some blocks in the mail out/mail back universe contained a significant number of non-city style address. They were converted to update/leave.

Table 4c: P-sample Preliminary Noninterview Rates in Occupied Housing Units by Type of Enumeration Area (Unweighted Data)	
Type of Enumeration Area	Percent Noninterview
Mail Out / Mail Back	
Update/Leave	
List/Enumerate	
Rural Update/Leave	
Urban Update/Leave	
Urban Update/Enumerate	
Additions to Address Listing	

The P-sample nonmatch rate is calculated by dividing the unweighted number of P-sample nonmatches by the unweighted P-sample total. This P-sample total does not include the people coded as removed. The E-sample nonmatch rate is also the unweighted number of E-sample nonmatches divided by the total unweighted E-sample. The percentage not matched for the P-sample and E-sample is in Table 5a by census region, in Table 5b by census regional office, in Table 5c by type of enumeration area, in Table 5d by respondent type and telephone interview, in Table 5e by sex, in Table 5f by age, in Table 5g by race, in Table 5h by Hispanic origin, in Table 5i by tenure, in Table 5j by size of metropolitan statistical area (MSA), and in Table 5k by mail return rates.

Table 5a: Percent Not Matched Before Follow-up by Census Region (Unweighted Data)		
Census Region	P-sample Percent Not Matched	E-sample Percent Not Matched
Northeast		
South		
Midwest		
West		
Total		

Table 5b: Percent Not Matched Before Follow-up by Census Regional Office (Unweighted Data)		
Census Regional Office	P-sample Percent Not Matched	E-sample Percent Not Matched
Boston		
New York		
Philadelphia		
Detroit		
Chicago		
Kansas City		
Seattle		
Charlotte		
Atlanta		
Dallas		
Denver		
Los Angeles		
Total		

Table 5c: Percent Not Matched Before Follow-up by Type of Enumeration Area (Unweighted Data)		
Type of Enumeration Area	P-sample Percent Not Matched	E-sample Percent Not Matched
Mail Out / Mail Back		
Update/Leave		
List/Enumerate		
Rural Update/Leave		
Urban Update/Leave		
Urban Update/Enumerate		
Additions to Address Listing		
Total		

Table 5d: Percent Not Matched Before Follow-up by Respondent Type and Telephone Interview (Unweighted Data)		
Respondent Type	P-sample Percent Not Matched	E-sample Percent Not Matched
Proxy		
Household Member		
Telephone Interview		
Person Interview		
Grand Total		

Table 5e: Percent Not Matched Before Follow-up by Sex (Unweighted Data)		
Sex	P-sample Percent Not Matched	E-sample Percent Not Matched
Male		
Female		
Blank		
Total		

Table 5f: Percent Not Matched Before Follow-up by Age (Unweighted Data)		
Age	P-sample Percent Not Matched	E-sample Percent Not Matched
Under 17		
18 to 29		
30 to 49		
50 to 65		
Over 65		
Blank		
Total		

Table 5g: Percent Not Matched Before Follow-up by Race (Unweighted Data)		
Race	P-sample Percent Not Matched	E-sample Percent Not Matched
White		
Black		
American Indian		
Asian		
Native Hawaiian and Pacific Islander		
Other Race		
Multiple Race		
Blank		
Total		

Table 5h: Percent Not Matched Before Follow-up by Hispanic Origin (Unweighted Data)		
Hispanic Origin	P-sample Percent Not Matched	E-sample Percent Not Matched
Hispanic		
Non-Hispanic		
Blank		
Total		

Table 5i: Percent Not Matched Before Follow-up by Tenure (Unweighted Data)		
Tenure	P-sample Percent Not Matched	E-sample Percent Not Matched
Owner		
Renter		
Blank		
Total		

The Metropolitan Statistical Area (MSA) variable is calculated from the Census 2000 unadjusted, total population count including group quarters, service based enumeration and remote Alaska people (Haines 2000). Table 5j contains three categories: large, medium, and small. Large consists of the ten largest MSAs. Medium consists of those with population size of at least 500,000 people. Small consists of those with a population size strictly less than 500,000 people. The last category is areas that are not in MSAs.

Table 5j: Percent Not Matched Before Follow-up by Size of Metropolitan Statistical Area (Unweighted Data)		
Metropolitan Statistical Area	P-sample Percent Not Matched	E-sample Percent Not Matched
Large		
Medium		
Small		
Non-MSA		
Total		

The percent not matched is in Table 5k for a cluster level variable measuring the proportion of occupied housing units in the mailback universe which returned a census questionnaire. Low return rate clusters are those clusters with return rates below 25 percent of occupied housing units. High return rate clusters are those clusters with return rates greater than or equal to 25 percent of occupied housing units. The last category is clusters where one or more of the housing units is not in the mail return universe.

Table 5k: Percent Not Matched Before Follow-up by Return Rates (Unweighted Data)		
Return Rates	P-sample Percent Not Matched	E-sample Percent Not Matched
High		
Low		
No mail return		
Total		

The Follow-up Interview

The person follow-up is conducted to gather additional information to accurately code the residence status of the nonmatched P-sample people and the enumeration status of the E-sample people. The following cases were sent to person follow-up:

- P-sample partial household nonmatches
- P-sample whole household nonmatches where the census enumerated different E-sample people (i.e., conflicting households or Smith/Jones cases)
- P-sample whole household nonmatches where the A.C.E. person interview was with a proxy respondent
- E-sample nonmatches
- Possible matches
- P-sample matches and nonmatches with unresolved residence status

The results of the follow-up interview are clerically entered into the matching software. Table 6 contains the results of the follow-up interviews for the P-sample nonmatches and possible matches. The P-sample people followed up are classified as

- Matched
- Not matched resident of the cluster on census day
- Unresolved residence or match status
- Nonresident of the cluster on census day and removed from the P-sample

Matched - The P-sample person was found in the census in the block cluster or in a surrounding block.

Not Matched resident of the cluster on census day - The P-sample nonmatch was not found in the census and should have been counted in the search area for this cluster.

Unresolved residence or match status - The person has unresolved residence status because the follow-up interview did not successfully collect the information required to accurately identify this person as a resident of the cluster on census day. In the case of possible matches, the interview did not accurately identify the match status of the people.

Removed from the P-sample - The P-sample person was not a resident of the housing unit on census day and is removed from the P-sample. These people are duplicates, fictitious, living in a P-sample housing unit that was listed in the cluster in error (i.e., P-sample geocoding error), or the P-sample person should have been counted at another residence on census day.

Table 6: Results of P-sample Follow-up for Nonmatches and Possible Matches

	Before Follow-up Match Code			
	Nonmatch		Possible Match	
After Follow-up Match Code	Unweighted People	Percent	Unweighted People	Percent
Matched				
Nonmatch Resident				
Unresolved				
Removed				
Total				

Table 7 contains the results of the follow-up interviews for the E-sample nonmatches and possible matches. The E-sample people followed up are classified as

- Matched
- Correctly enumerated
- Erroneously enumerated
- Unresolved

Matched - The P-sample and E-sample people refer to the same person.

Correctly enumerated - The E-sample nonmatch was correctly enumerated in the census.

Erroneously enumerated - The E-sample nonmatch was erroneously enumerated in the census, because the person should have been counted at another residence on census day, was fictitious, had insufficient information for matching and follow-up, was duplicated, or lived in a household

that was a geocoding error.

Unresolved - The follow-up interview for the census nonmatch was not successful.

Table 7: Results of E-sample Follow-up for Nonmatches and Possible Matches				
	Before Follow-up Match Code			
	Nonmatch		Possible Match	
After Follow-up Match Code	Unweighted People	Percent	Unweighted People	Percent
Matched				
Correctly Enumerated				
Erroneously Enumerated				
Unresolved				
Total				

After Follow-up Match Results

The final P-sample results are in Tables 8 and 9. The P-sample people have been classified as matched, not matched, unresolved match status, and removed in Table 8 and also tabulated as resident, nonresident, and unresolved residence status in Table 9.

The P-sample match status is defined as

- matched
- not matched
- unresolved match status
- removed from the P-sample

Matched - The P-sample was found in the cluster or in the surrounding block in either a housing unit or in group quarters.

Not matched - The P-sample person was not found in the search area. If the nonmatch was sent to follow-up, the person was confirmed to be a resident of the cluster on census day. If the nonmatch was not sent for a follow-up interview, a household member identified the person as a resident of the housing unit during the original A.C.E. interview.

Unresolved match status - The match status is unresolved for possible matches with unsuccessful follow-up interviews and for P-sample people with insufficient information for matching and follow-up.

Removed from the P-sample - People are removed from the P-sample when they are fictitious, duplicates, geocoding errors, or not residents of the housing unit on census day.

The P-sample residence status is defined as

- resident
- nonresident
- unresolved residence status

Resident - The P-sample matched or not matched person is a resident of the housing unit on census day.

Nonresident - P-sample people are nonresidents of the cluster when they are fictitious, duplicates, geocoding errors, or should not have been included as a resident of the housing unit on census day. Nonresidents are removed from the P-sample.

Unresolved Residence Status - A matched or not matched P-sample person has unresolved residence status when the follow-up interview did not successfully determine the person's residence on census day. The residence status of the possible match is unresolved when the follow-up interview was not successful. The residence status is also imputed when the P-sample person had insufficient information for matching.

The final E-sample results are in Table 10. The E-sample people have been classified as correctly or erroneously enumerated and enumeration status of unresolved. These are the unweighted match results that go to imputation and estimation.

The E-sample enumeration status is defined as

- correctly enumerated
- erroneously enumerated
- unresolved enumeration status

Correctly Enumerated - E-sample people are correctly enumerated when they are matched to the P-sample or when they have been followed up and they should have been enumerated in this cluster.

Erroneously Enumerated - E-sample people are erroneously enumerated when they have another residence where they should be counted on census day, are fictitious, are duplicated, lived in a housing unit that was a geocoding error, or have insufficient information for matching and follow-up.

Unresolved Enumeration Status - E-sample people have unresolved enumeration status when the follow-up interview was unsuccessful. The E-sample person may have been followed up to obtain information about the E-sample nonmatch, possible match, matched person with unresolved residence status, or geographic work to obtain the location of the housing unit.

Table 8: National P-Sample Match Status After Follow-up		
P-sample Match Status	Unweighted People	Percent
Matched		
Not Matched		
Unresolved		
Removed		
Total		

Table 9: National P-Sample Residence Status After Follow-up		
P-sample Residence Status	Unweighted People	Percent
Resident		
Nonresident		
Unresolved		
Total		

Table 10: National E-Sample Matching After Follow-up		
E-sample Enumeration Status	Unweighted People	Percent
Correctly Enumerated		
Erroneously Enumerated		
Unresolved		
Total		

The percent P-sample not matched and E-sample erroneous enumeration is contained in Table 11a by census region, in Table 11b by census regional office, in Table 11c by type of enumeration area, in Table 11d by respondent type and telephone interview, in Table 11e by sex, in Table 11f by age, in Table 11g by race, in Table 11h by Hispanic origin, in Table 11i by

tenure, in Table 11j by size of metropolitan statistical area (MSA), and in Table 11k by mail return rates. The percent P-sample not matched is one hundred times the nonmatch rate.

$$\text{Nonmatch Rate} = \frac{\text{Not Matched}}{\text{Matched + Not Matched}}$$

The percent E-sample erroneous enumeration is one hundred times the erroneous enumeration rate.

$$\text{Erroneous Enumeration rate} = \frac{\text{Erroneous Enumeration}}{\text{Correct Enumeration + Erroneous Enumeration}}$$

Both percentages are of unweighted resolved people in Tables 11a through 11k. The weighting and imputation process happens after the matching is completed.

Table 11a: Percent Not Matched and Erroneously Enumerated After Follow-up by Census Region Before Weighting and Imputation

Census Region	P-sample Percent Not Matched	E-sample Percent Erroneous Enumeration
Northeast		
South		
Midwest		
West		
Total		

Table 11b: Percent Not Matched and Erroneously Enumerated After Follow-up by Census Regional Office Before Weighting and Imputation		
Census Regional Office	P-sample Percent Not Matched	E-sample Percent Erroneous Enumeration
Boston		
New York		
Philadelphia		
Detroit		
Chicago		
Kansas City		
Seattle		
Charlotte		
Atlanta		
Dallas		
Denver		
Los Angeles		
Total		

Table 11c: Percent Not Matched and Erroneously Enumerated After Follow-up by Type of Enumeration Area Before Weighting and Imputation		
Type of Enumeration Area	P-sample Percent Not Matched	E-sample Percent Erroneous Enumeration
Mail Out / Mail Back		
Update/Leave		
List/Enumerate		
Rural Update/Leave		
Urban Update/Leave		
Urban Update/ Enumerate		
Additions to Address Listing		
Total		

Table 11d: Percent Not Matched and Erroneously Enumerated After Follow-up by Respondent Type and Telephone Interview Before Weighting and Imputation		
Respondent Type	P-sample Percent Not Matched	E-sample Percent Erroneous Enumeration
Proxy		
Household Member		
Telephone Interview		
Person Interview		
Grand Total		

Table 11e: Percent Not Matched and Erroneously Enumerated After Follow-up by Sex Before Weighting and Imputation		
Sex	P-sample Percent Not Matched	E-sample Percent Erroneous Enumeration
Male		
Female		
Blank		
Total		

Table 11f: Percent Not Matched and Erroneously Enumerated After Follow-up by Age Before Weighting and Imputation		
Age	P-sample Percent Not Matched	E-sample Percent Erroneous Enumeration
Under 17		
18 to 29		
30 to 49		
50 to 65		
Over 65		
Blank		
Total		

Table 11g: Percent Not Matched and Erroneously Enumerated After Follow-up by Race Before Weighting and Imputation		
Race	P-sample Percent Not Matched	E-sample Percent Erroneous Enumeration
White		
Black		
American Indian		
Asian		
Native Hawaiian and Pacific Islander		
Other Race		
Multiple Race		
Blank		
Total		

Table 11h: Percent Not Matched and Erroneously Enumerated After Follow-up by Hispanic Origin Before Weighting and Imputation		
Hispanic Origin	P-sample Percent Not Matched	E-sample Percent Not Matched
Hispanic		
Non-Hispanic		
Blank		
Total		

Table 11i: Percent Not Matched and Erroneously Enumerated After Follow-up by Tenure Before Weighting and Imputation		
Tenure	P-sample Percent Not Matched	E-sample Percent Not Matched
Owner		
Renter		
Blank		
Total		

Table 11j: Percent Not Matched and Erroneously Enumerated After Follow-up by Size of Metropolitan Statistical Area Before Weighting and Imputation		
Metropolitan Statistical Area	P-sample Percent Not Matched	E-sample Percent Not Matched
Large		
Medium		
Small		
Non-MSA		
Total		

Table 11k: Percent Not Matched and Erroneously Enumerated After Follow-up by Return Rates Before Weighting and Imputation		
Return Rates	P-sample Percent Not Matched	E-sample Percent Not Matched
High		
Low		
No mail return		
Total		

The unresolved codes for the P-sample are either unresolved match status or unresolved residence status. Cases with unresolved residence status are: matches with unsuccessful follow-

up interview and nonmatches with unsuccessful follow-up interview. Cases with both unresolved match and unresolved residence status are: possible matches with unsuccessful follow-up interview and the P-sample people with insufficient information for matching and follow-up. Tables 12a through 12k contain the percent of the total P-sample with unresolved match or residence status. All people with unresolved match status also have unresolved residence status, so the percentage with unresolved match status or unresolved residence status is the percentage of unresolved residence status. The percentage with unresolved match or residence status is one hundred times the rate of unresolved match or residence status.

$$\text{Rate of Unresolved Match or Residence Status} = \frac{\text{People with Unresolved Residence Status}}{\text{Total P-sample People}}$$

Cases with unresolved enumeration status for the E-sample are the ones with unsuccessful follow-up interview. Tables 12a through 12k also contain the percent of the total E-sample with unresolved enumeration status. The percentage with unresolved enumeration status is one hundred times the rate of unresolved enumeration status.

$$\text{Rate of Unresolved Enumeration Status} = \frac{\text{People with Unresolved Enumeration Status}}{\text{Total E-sample People}}$$

The percentage unresolved for the P-sample and E-sample is in Table 12a by census region, in Table 12b by census regional office, in Table 12c by type of enumeration area, in Table 12d by respondent type and telephone interview, in Table 12e by sex, in Table 12f by age, in Table 12g by race, in Table 12h by Hispanic Origin, in Table 12i by tenure, in Table 12j by size of metropolitan statistical area (MSA), and in Table 12k by mail return rates.

Table 12a: Percent Unresolved After Follow-up by Census Region Before Weighting and Imputation

Census Region	Percent P-sample Unresolved	Percent E-sample Unresolved
Northeast		
South		
Midwest		
West		
Total		

Table 12b: Percent Unresolved After Follow-up by Census Regional Office Before Weighting and Imputation		
Census Regional Office	Percent P-sample Unresolved	Percent E-sample Unresolved
Boston		
New York		
Philadelphia		
Detroit		
Chicago		
Kansas City		
Seattle		
Charlotte		
Atlanta		
Dallas		
Denver		
Los Angeles		
Total		

Table 12c: Percent Unresolved After Follow-up by Type of Enumeration Area Before Weighting and Imputation		
Type of Enumeration Area	Percent P-sample Unresolved	Percent E-sample Unresolved
Mail Out / Mail Back		
Update/Leave		
List/Enumerate		
Rural Update/Leave		
Urban Update/Leave		
Urban Update/ Enumerate		
Additions to Address Listing		
Total		

Table 12d: Percent Unresolved After Follow-up by Respondent Type and Telephone Interview Before Weighting and Imputation		
Respondent Type	Percent P-sample Unresolved	Percent E-sample Unresolved
Proxy		
Household Member		
Telephone Interview		
Person Interview		
Grand Total		

Table 12e: Percent Unresolved After Follow-up by Sex Before Weighting and Imputation		
Sex	Percent P-sample Unresolved	Percent E-sample Unresolved
Male		
Female		
Blank		
Total		

Table 12f: Percent Unresolved After Follow-up by Age Before Weighting and Imputation		
Age	Percent P-sample Unresolved	Percent E-sample Unresolved
Under 17		
18 to 29		
30 to 49		
50 to 65		
Over 65		
Blank		
Total		

Table 12g: Percent Unresolved After Follow-up by Race Before Weighting and Imputation		
Race/Hispanic	Percent P-sample Unresolved	Percent E-sample Unresolved
White		
Black		
American Indian		
Asian		
Native Hawaiian and Pacific Islander		
Other Race		
Multiple Race		
Blank		
Total		

Table 12h: Percent Unresolved After Follow-up by Hispanic Origin Before Weighting and Imputation (Unweighted Data)		
Hispanic Origin	P-sample Percent Unresolved	E-sample Percent Unresolved
Hispanic		
Non-Hispanic		
Blank		
Total		

Table 12i: Percent Unresolved After Follow-up by Tenure Before Weighting and Imputation		
Tenure	Percent P-sample Unresolved	Percent E-sample Unresolved
Owner		
Renter		
Blank		
Total		

Table 12j: Percent Unresolved After Follow-up by Metropolitan Statistical Area Before Weighting and Imputation		
Metropolitan Statistical Area	Percent P-sample Unresolved	Percent E-sample Unresolved
Large		
Medium		
Small		
Non-MSA		
Total		

Table 12k: Percent Unresolved Before Follow-up by Mail Return Rates Before Weighting and Imputation		
Mail Return Rates	Percent P-sample Unresolved	Percent E-sample Unresolved
High		
Low		
No mail return		
Total		

Erroneously enumerated - The categories are people with insufficient information for matching and follow-up, duplicates, fictitious, geocoding errors, and people who should have been enumerated at another residence on census day.

- The E-sample people with insufficient information for matching and follow-up are ones who are data defined, but do not contain full name and at least two characteristics.
- The E-sample people enumerated more than once are coded as duplicates.
- The fictitious people are ones where we found notes on the census image identifying the person as not a real person such as a dog or other pet or they were identified as not existing in this cluster during the follow-up interview.
- Census people in housing units identified as geocoding errors during the housing unit follow-up are coded as erroneously enumerated because of geocoding error.
- The E-sample person should have been counted at another residence on census day.

The percentages of each type of erroneous enumeration in Tables 13a and 13b are based on the E-sample people with a resolved enumeration status. Note that the percentage of each type of erroneous enumeration is one hundred times the rate of each type of erroneous enumeration.

$$\text{Rate of Type of Erroneous Enumeration} = \frac{\text{Type of Erroneous Enumeration}}{\text{Correct Enumeration + Erroneous Enumeration}}$$

Table 13a has the type of erroneous enumeration as a percent of the total E-sample resolved cases by type of enumeration area and Table 13b has this percent by race.

Table 13a: Type of Erroneous Enumeration After Follow-up Before Weighting and Imputation

	Percent of E-sample Resolved		
Erroneous Enumeration Codes	Mail Out/ Mail Back	All Other Types of Enumeration Areas	Total
Insufficient Information			
Duplicate			
Fictitious			
Geocoding Error			
Other Residence			
Total			

Table 13b: Type of Erroneous Enumeration After Follow-up by Race Before Weighting and Imputation								
	Percent of Resolved E-sample							
Erroneous Enumeration Codes	White	Black	American Indian	Asian	Native Hawaiian, Pacific Islander	Other Race	Multiple Race	Blank
Insufficient Information								
Duplicate								
Fictitious								
Geocoding Error								
Other Residence								
Total								

Final Census Day Interview Outcome

The final census day outcome codes are in Tables 14 by interview mode and total. Changes as a result of the follow-up interview are

- Whole households of P-sample people who said they lived elsewhere on census day are converted to a noninterviews.
- Whole households who lived in group quarters on census day or should have been enumerated at another residence are converted to vacant.

Table 14: Final Census Day Estimation Outcome Codes for P-sample Housing Units by Telephone Interview (Unweighted Data)						
	Telephone Interview		Person Interview		Total	
Outcome Code	Housing Units	Percent	Housing Units	Percent	Housing Units	Percent
Complete interview with a household member						
Complete interview with a proxy respondent						
Partial interview						
No census day residents - household converted to noninterview						
Field noninterview						
All people have insufficient information for matching and follow-up						
No census day residents - Converted to vacant						
Vacant on census day						
Not a housing unit on census day						
Total						

The census day noninterview rates are recalculated to reflect changes due to coding in after follow-up matching. The final noninterview rates are in Table 15a by telephone versus person interview, Table 15b by census regional offices, Table 15c by census region, and Table 15d by type of enumeration area.

Table 15a: P-sample Noninterview Rates in Occupied Housing Units by Interview Mode (Unweighted Data)	
Interview Mode	Percent Noninterview
Telephone	
Personal	

Table 15b: P-sample Noninterview Rates in Occupied Housing Units by Census Regional Office (Unweighted Data)	
Census Regional Office	Percent Noninterview
Boston	
New York	
Philadelphia	
Detroit	
Chicago	
Kansas City	
Seattle	
Charlotte	
Atlanta	
Dallas	
Denver	
Los Angeles	
Total	

Table 15c: P-sample Noninterview Rates in Occupied Housing Units by Census Region (Unweighted Data)	
Census Region	Percent Noninterview
Northeast	
South	
Midwest	
West	

Table 15d: P-sample Noninterview Rates in Occupied Housing Units by Type of Enumeration Area (Unweighted Data)	
Type of Enumeration Area	Percent Noninterview
Mail Out / Mail Back	
Update/Leave	
List/Enumerate	
Rural Update/Leave	
Urban Update/Leave	
Urban Update/Enumerate	
Additions to Address Listing	

Final Interview Day Interview Outcome

The final interview day outcome codes are in Table 16 by interview mode. The interview outcome as of interview day is for the nonmovers and the inmovers. Changes as a result of the follow-up interview are in whole households of nonmovers who said they lived elsewhere, in group quarters, or have another residence where they should have been counted on census day are converted to noninterviews.

Table 16: Final Interview Day Estimation Outcome Codes for P-sample Housing Units by Interview Mode
(Unweighted Data)

	Telephone Interview		Person Interview		Total	
Outcome Code	Housing Units	Percent	Housing Units	Percent	Housing Units	Percent
Complete interview with a household member						
Complete interview with a proxy respondent						
Partial interview						
No census day residents - household converted to noninterview						
Field noninterview						
All people have insufficient information for matching and follow-up						
Vacant on census day						
Not a housing unit on census day						
Total						

The interview day noninterview rates are recalculated to reflect changes due to coding in after follow-up matching. The final noninterview rates are in Table 17a by telephone versus person interview, in Table 17b for census regional offices, in Table 17c by census region, and in Table 17d by type of enumeration area.

Table 17a: P-sample Interview Day Noninterview Rates in Occupied Housing Units by Interview Mode (Unweighted Data)	
Interview Mode	Percent Noninterview
Telephone	
Personal	

Table 17b: P-sample Interview Day Noninterview Rates in Occupied Housing Units by Census Regional Office (Unweighted Data)	
Census Regional Office	Percent Noninterview
Boston	
New York	
Philadelphia	
Detroit	
Chicago	
Kansas City	
Seattle	
Charlotte	
Atlanta	
Dallas	
Denver	
Los Angeles	
Total	

Table 17c: P-sample Interview Day Noninterview Rates in Occupied Housing Units by Census Region (Unweighted Data)	
Census Region	Percent Noninterview
Northeast	
South	
Midwest	
West	

Table 17d: P-sample Interview Day Noninterview Rates in Occupied Housing Units by Type of Enumeration Area (Unweighted Data)	
Type of Enumeration Area	Percent Noninterview
Mail Out / Mail Back	
Update/Leave	
List/Enumerate	
Rural Update/Leave	
Urban Update/Leave	
Urban Update/Enumerate	
Additions to Address Listing	

Outlier Journals

Journals describing the cluster are being written for the clusters with the most influence in the dual system estimator to provide documentation for clusters with high rates of weighted P-sample nonmatches and/or E-sample erroneous enumerations.

Quality Assurance of the Clerical Person Matching Operation

The Person Matching for the Accuracy and Coverage Evaluation took place in two phases, Computer Matching and Clerical Matching. The computer matching used the Statistical Research Division's matching software to match the P-sample to the census. After the computer matching, P-sample and E-sample people who possibly matched or who did not match were clerically reviewed in the Clerical Matching Operation. This document reports on the results of the Quality Assurance for the Clerical Matching.

The Clerical Person Matching in the National Processing Center (NPC) was conducted by three levels of matchers - clerks, technicians, and analysts. There were xxx clerks, xx technicians, and xx analysts. The Quality Assurance (QA) plan for the clerical person matching operation consists of a three-tiered, dependent review of the work done by the clerical matchers. This review was accomplished by a dependent rework of samples of each person's work, a process that should identify random matching errors. This plan was designed so that each of the matching levels improves on the previous level. The clerks match what the computer matcher could not. The technicians work on any cases the clerks could not resolve and perform the quality assurance on the clerks. Then the analysts finish any cases the technicians could not resolve and perform the quality assurance on the technicians. Clusters with match results that require a second opinion are sent to the higher stage.

The before followup (BFU) QA plan is primarily aimed at finding errors that could have prevented people who should have been followed up from being assigned to the followup operation. The after followup (AFU) QA plan was aimed at finding errors that could have caused people to have been incorrectly classified as either correctly or erroneously enumerated, as well as errors that could have caused people to incorrectly drop out of the P sample.

The QA plan was designed to control the quality of the clerical matching by targeting both records that require a higher level of review and individual matchers who require more consistent review. The Person Matching Review and Coding System (PERMaRCS) began by monitoring the work of all matchers to target matchers who required a more consistent review. All users (clerks and technicians) began with 100% of their work being reviewed and were then approved for having a sample of their work reviewed only after satisfactorily completing a given amount of work. PERMaRCS monitored each clerk's and technician's matching results through the entire matching process by counting significant changes[5] of codes. Each time a matcher worked a predetermined number of records, the system reassessed that matcher's sampling status. Based

[5]Most changes were considered significant, however certain codes could be entered by the clerk or technician to flag a record for review by the higher level tech or analyst. In this case, any code change was not considered significant. In other cases the distinction between the codes and the resulting damage if incorrectly coding a record with these codes was considered insignificant and the change was not counted.

on the counts of changes, some users were placed back into 100% review while others were put into sampling mode. In this way, the sampling QA decision was periodically revisited for each matcher.

To target records that required a higher level of review, the system allowed users to flag difficult records for the next level of review. The system also checked for certain predefined situations and if present, the cluster or batch was automatically sent to a higher level matcher for review. For example, if a technician changed more than half of a clerk's codes, an analyst reviewed all of the coding for that clerk and that technician.

Additionally, the matching software was designed to continuously check the matcher's work to minimize many kinds of mistakes, such as assigning an invalid match code or leaving some records uncoded.

Assumptions of the QA Plan

The QA Plan has three major assumptions:

- The change rate is an overestimate of the true error rate.

Because matching person data is subjective, individual code changes do not always indicate errors. The QA plan assumes a negative correlation between a matcher whose coding is frequently changed and quality of coding, but not a one-to-one correspondence between code changes and errors.

In certain situations, records went for a higher level review but were still considered to be in the out-of-sample workload. Clusters or batches of the work went to the tech or analyst either for a complete review (for matchers in 100% review or in samples selected for review), or because the cluster or batch contained a situation flagged by the system for higher review. Records in this latter category were usually indistinguishable from records that did not go to a higher review, and were therefore considered out-of-sample. Such records are considered to be worked by a clerk or tech, when in reality these were reviewed by an analyst. Consequently, in estimating the remaining error, the change rate was applied to an inflated total out-of-sample count that included these records.

- The QA plan measures the quality of the clerks' and techs' work, not the analysts.

There is no QA of the analysts' work. Due to their extensive training and specific knowledge of the task, analysts are assumed to have no errors. Differences between analysts' codes are likely to be a result of the subjective nature of matching and cannot be distinguished from inaccuracies.

- Each stage is considered individually and the last two stages of AFU are disregarded for QA purposes.

The rules for coding records in BFU are different than the rules for AFU and consequently the types of errors are different. Therefore the AFU results cannot be used to measure errors in BFU. Additionally, within the AFU stage, the rules and types of errors in the clerk, tech, and analyst stages are different from the coding rules and types of errors in cluster review and outlier stages. The first three stages review the records individually (in batches) while the last two stages reconstruct the cluster for a cluster-level review. Therefore a change made to a record in the last two stages of AFU does not indicate an error in the first three stages of AFU. These changes are the result of additional information available during cluster review.

Because the last two stages do not have QA, any code changes in these stages are disregarded for the outgoing quality calculation.

QA Results of the Clerical Matching Operation

The QA results of the clerical matching provide individual and overall change rates for clerks and techs in the BFU and AFU stages. These calculations can then be used to estimate the number of defects in records worked by clerks and techs that did not get a higher review. For each stage of matching, the records are partitioned into the highest stage of review (BFU clerk, BFU tech, BFU analyst, AFU clerk, AFU tech, AFU analyst), and a resulting outgoing quality by stage is determined for the records completed in this stage (and not reviewed in any higher stage).

- BFU Tech Change Rate

The BFU Tech individual change rate is derived for each tech by looking at the records worked by that tech. For each tech the number of records changed by the analysts is divided by the number of records checked by the analysts. This is considered to be the individual BFU Tech change rate. To estimate the remaining errors in this tech's work, this rate is multiplied by the number of records worked by this tech that underwent no higher review. An overall BFU Tech change rate will be calculated using a weighted average of all the individual tech change rates. The overall quality of the BFU Tech stage is calculated as shown below.

- BFU Tech individual change rate, $errT_i = chgsAT_i \div chkAT_i$
  where
  $chgsAT_i$ are the records worked by tech T_i and changed by an analyst, and
  $chkAT_i$ are the records worked by tech T_i and checked by an analyst.
- BFU Tech overall change rate, $BFT\_err = \left(\sum_i^n errT_i * recsT_i\right) \div \left(\sum_i^n recsT_i\right)$
  where
  $errT_i$ is the BFU Tech individual change rate for tech T_i, and
  $recsT_i$ is the number of records worked by tech T_i and not reviewed by an analyst.

- Overall quality of the BFU Tech stage = (1 - *BFT_err*)*100

- BFU Clerk Change Rate

The BFU Clerk individual change rate will be calculated by determining the number of records worked by a given clerk that were changed in higher stages of review and then dividing this number by the total number of records worked by the clerk that were checked in a higher stage. This calculation will be done in two parts to distinguish those records worked by the clerk and reviewed only by a technician from records worked by the clerk and reviewed by an analyst. This distinction is made because the records reviewed only by a tech are assumed to contain some errors due to the technician. Records reviewed only by techs will be assigned an adjusted overall tech change rate and this rate will be used to augment the individual clerk change rate to account for the tech's potential missed errors. Since this results in estimating more changes in the clerks' work than were actually counted, an adjusted rate is used to exclude those cases where the tech's code was changed, but the clerk properly coded the record. To estimate the remaining errors in a clerk's work, this rate is multiplied by the number of records worked by the clerk with no higher review. An overall BFU Clerk change rate will be calculated using a weighted average of all the individual clerk change rates. The overall quality of the BFU Clerk stage is calculated as shown below.

- Adjusted BFU Tech overall change rate, *BFT_err_adj*

$$=\left(\sum_{i}^{n}\left(\frac{chgsAT_i - corrCT_i}{chkAT_i}\right) * recsT_i\right) \div \left(\sum_{i}^{n} recsT_i\right)$$

where
$chgsAT_i$ is the number of records worked by tech T_i and changed by an analyst,
corrCTi is the number of those records reviewed by tech T_i and changed by an analyst where a clerk coded the record correctly,
$chkAT_i$ is the number of tech T_i's records checked by an analyst, and
$recsT_i$ is the number of records worked by tech T_i and not reviewed by an analyst.

- BFU Clerk individual change rate, $errC_i$ =

$$\left[chgsAC_i + chgsTC_i + \left(BFT\_err\_adj * chkTC_i\right)\right] \div \left[chkAC_i + chkTC_i\right]$$

where
$chgsAC_i$ are the records worked by clerk C_i and changed by an analyst,
$chgsTC_i$ are the records worked by clerk C_i and changed by a tech,
BFT_err_adj is the Adjusted BFU Tech overall change rate,
$chkTC_i$ is the number of clerk C_i's records checked only by techs, *BFT_err_adj* is multiplied by $chkTC_i$ to estimate the potential errors the techs may have missed in this clerk's work, and
$chkAC_i$ is the number of records worked by clerk C_i and checked by an analyst.

- BFU Clerk overall change rate, BFC_err =

$$\left(\sum_{i}^{n} errC_i * recsC_i\right) \div \left(\sum_{i}^{n} recsC_i\right)$$

where
$errC_i$ is the BFU Clerk individual change rate for clerk C_i, and
$recsC_i$ is the number of records coded by clerk C_i and not reviewed by a tech or an analyst.

- Overall quality of the BFU Clerk stage = (1 - BFC_err)*100

- AFU Tech Change Rate

The AFU Tech change rates, both individual and overall, and the overall quality in the AFU Tech stage, are calculated the same way as the BFU Tech change rates only using records from the AFU Tech and AFA stages.

- AFU Clerk Change Rate

The AFU Clerk change rates, both individual and overall, and the overall quality in the AFU Tech stage, are calculated the same way as the BFU Clerk change rates, again considering records reviewed by technicians differently than those reviewed by analysts.

Table 18 shows the coding changes by level of highest review. Records worked in the clerical matching operation only appear in one row, that is, if a clerk and tech both worked the record in BFU, this record appears only in the BFU Tech row.

Table 18: Coding Changes by Level of Highest Review

Stage	Tech Changes	Total records reviewed by Techs	Analyst Changes	Total records reviewed by Analysts
BFU Clerk				
BFU Tech	n/a	n/a		
AFU Clerk				
AFU Tech	n/a	n/a		
Total (n)				

Table 19 shows the overall tech and clerk change rates by stage.

Table 19: Overall Change Rate by Stage	
BFU Clerk change rate	
BFU Tech change rate	
BFU Analyst change rate	0.0%
AFU Clerk change rate	
AFU Tech change rate	
AFU Analyst change rate	0.0%

Table 20 shows the number of records which received a final code at the given stage.

Table 20: Records Completed by Stage							
	BFU Clerk	BFU Tech	BFU Analyst	AFU Clerk	AFU Tech	AFU Analyst	Total
Records completed							
Percentage of total workload completed							100%

The estimated resulting outgoing quality by stage is presented in Table 21. Individual estimates of remaining error rates for techs and clerks in BFU and AFU are presented in Appendix B.

Table 21: Percentage of Resulting Outgoing Quality by Stage	
Stage	Resulting Outgoing Quality Rate
BFU Clerk	
BFU Tech	
BFU Analyst	100%
AFU Clerk	
AFU Tech	
AFU Analyst	100%

Limitations

The QA results presented in this document report only on the quality of the clerical matching stages in BFU and the first three stages of AFU. However, xx% of the total person matching workload was matched in the Computer Matching Phase by the Statistical Research Division's computer matching software. Computer matching is very conservative. Numerous studies over the years have shown that this operation is virtually error free (insignificant numbers of false matches).

Person Followup Quality Assurance

The quality assurance plan for PFU involves two parts: a data edit to ensure completeness of the PFU form and a recontact of the respondent to detect falsification.

After the PFU form is completed by the interviewer, the supervisor (a crew leader) reviews the PFU form for legibility and completeness. If any part of the form is illegible, then the crew leader contacts the interviewer or respondent to determine the proper answer. If any skip patterns are not followed or any section of the form is incorrectly left blank, then the crew leader contacts the interviewer or sends the questionnaire back to the field for clarification. After the crew leader edit, the form is sent to the A.C.E. Regional Office where staff reviews the form for completeness and legibility again.

To detect falsification, a sample of forms is selected for QA. Three types of cases are sent to PFU QA:

1. A random sample of 1-in-20
2. The first eligible form for each interviewer that is returned from the field

3. Supervisor-selected cases if an office supervisor suspects falsification

An office QA checker attempts to contact the respondent from the PFU form by telephone. The QA checker asks if the respondent was recently contacted by a Census Bureau employee. If the respondent was contacted, then the case passes QA. If the respondent was not contacted, the QA checker conducts the interview with the respondent. If the respondent cannot be contacted the case is sent back to the field to attempt to contact the respondent. Again, the QA checker attempts to determine if the respondent was contacted for the PFU interview. The field QA checker follows the same procedures as the office QA checker.

For PFU QA, the respondent recontact is limited to certain cases. Only completed cases with one respondent are eligible to be recontacted. In addition, a total of 6 days is allowed to attempt to recontact the respondent: 3 days by telephone and 3 days by personal visit.

Person Followup Quality Assurance Results

Table 22 contains the QA selection status of PFU cases. Table 23 contains the workload for PFU QA by A.C.E. Regional Office. Table 24 contains the outcome of cases in PFU QA. Table 25 contains the mode of PFU QA recontact. Table 26 contains the QA outcome by mode of contact.

Table 22: QA Selection Status of PFU Cases

QA Status	Number of Cases	Percentage of PFU Cases
Selected		
Eligible		
Ineligible		
Not Selected		
Grand Total		

Table 23: Workload for PFU QA by A.C.E. Regional Office			
A.C.E. Regional Office	Cases in PFU QA	Percent of Cases in PFU QA	Percent of Cases in PFU
Boston			
New York			
Philadelphia			
Detroit			
Chicago			
Kansas City			
Seattle			
Charlotte			
Atlanta			
Dallas			
Denver			
Los Angeles			
Totals			

Table 24: QA Outcome of Cases in PFU QA		
Outcome	Number of Cases	Percentage of Cases in PFU QA
Pass		
Fail[6]		
Noninterview		
Totals		

[6]For PFU QA, a case is determined to fail if a replacement interview is taken (i.e., the respondent has said that he or she was not contacted for the original interview).

Table 25: Mode of PFU QA Recontact		
Mode	Number of Cases	Percentage of Cases in PFU QA
Telephone		
Personal Visit		
Totals		

Table 26: PFU QA Outcome by Mode of Contact		
Mode/Outcome	Number of Cases	Percentage of Cases in PFU QA
Telephone		
Pass		
Fail		
Noninterview		
Personal Visit		
Pass		
Fail		
Noninterview		
Grand Totals		

References

Byrne, R. (May 2000), "MaRCS Specifications for Quality Assurance (QA) for Person Matching", DSSD Census 2000 Procedures and Operations Memorandum Series Chapter S-QA-13.

Byrne, R. (May 2000), "Quality Assurance for A.C.E. Person Follow-up", DSSD Census 2000 Procedures and Operations Memorandum Series Chapter S-QA-14.

Childers, D. (September 2000), "Accuracy and Coverage Evaluation: The Design Document", Census 2000 Procedures and Operations Memorandum Series, Chapter S-DT-1.

Davis, M. (July 1991), "Preliminary Final Report for PES Evaluation Project P7: Estimates of P-sample Clerical Matching Error from a Rematching Evaluation", 1990 Coverage Studies and

Evaluation Memorandum Series # H-2.

Davis, M. (July 1991a), "Preliminary Final Report for PES Evaluation Project P10: Measurement of the Census Erroneous Enumerations–Clerical Error made in the Assignment of Enumeration Status", 1990 Coverage Studies and Evaluation Memorandum Series # L-2.

Feindt, P. and Byrne, R. (September 2000), "Accuracy and Coverage Evaluation Survey: Person Interviewing Results", Census 2000 Procedures and Operations Memorandum Series, Chapter B.

Haines, D. (May 2000), "Accuracy and Coverage Evaluation Survey: Computer Specification for Person Dual System Estimation (U. S.) ", DSSD Census 2000 Procedures and Operations Memorandum Series Chapter Q-29.

West, K. (July 1991b), "Final Report for PES Evaluation Project P4: Quality of Reported Census Day Address – Evaluation Follow-up", 1990 Coverage Studies and Evaluation Memorandum Series # D-2.

"Assessment on Accuracy of Adjusted Versus Unadjusted 1990 Census Base for Use in Intercensal Estimates". Report of the Committee on Adjustment of Postcensal Estimates, United States Bureau of the Census, Washington, D. C., August 7, 1992.

"Additional Research on Accuracy of Adjusted Versus Unadjusted 1990 Census Base for Use in Intercensal Estimates". Report of the Committee on Adjustment of Postcensal Estimates, United States Bureau of the Census, Washington, D. C., November 25, 1992.